UNIQUE REPRESENTATIONS OF REAL NUMBERS IN NON-INTEGER BASES

PAUL GLENDINNING AND NIKITA SIDOROV

1. Introduction

Problems related to the expansions of real numbers in non-integer bases have been systematically studied since the late 1950's, starting with the seminal works by Rényi [16] and Parry [15]. The original approach is based on a specific algorithm for choosing "digits" (e.g. the greedy expansions). This usually leads to the set of sequences of digits for all possible real numbers in question (for instance, non-negative or belonging to a given interval) which, unlike the classical d-adic case, is not a Cartesian product but has a complicated structure.

However, in the 1990's a group of Hungarian mathematicians led by Paul Erdös began to investigate 0-1 sequences that provide *unique* representations of reals [6, 7, 8]. The present paper continues this line of research. Our set-up is as follows.

Let $q \in (1, 2)$ be our parameter and $\Sigma = \prod_1^\infty \{0, 1\}$; we consider those x which have unique expansions in base q of the form

$$(1) \qquad x = \sum_{n=1}^\infty \varepsilon_n q^{-n}$$

with $\varepsilon = (\varepsilon_n)_1^\infty \in \Sigma$. The special case of $x = 1$ has received a lot of attention (see [6, 7, 8]), and recently Komornik and Loreti [12] have shown that there is a smallest $q \in (1, 2)$, denoted by q_c, for which there is only one representation of $x = 1$ in this form. Moreover, they show that q_c is the unique solution of the equation

$$\sum_1^\infty \mathfrak{m}_n x^{-n+1} = 1,$$

where $\mathfrak{m} = (\mathfrak{m}_n)_1^\infty$ is the famous Thue-Morse sequence

$$(2) \qquad \mathfrak{m} = 0110\ 1001\ 1001\ 0110\ 1001\ 0110\ 0110\ 1001 \ldots$$

(see, e.g., [4]). From this they obtain the numerical value

$$q_c = 1.787231650 \ldots$$

Recently, Allouche and Cosnard [2] have shown that q_c is transcendental (see also [3]).

In this paper we return to the case of general x in (1). Given $q \in (1,2)$ we aim to describe some properties of the set of all $x \geq 0$ for which the representation (1) with $\varepsilon \in \Sigma$ exists and is unique.

2. Formulation of the main result

Our first remark consists in the observation that since $\varepsilon \in \Sigma$, x must belong to $[0, 1/(q-1)]$ and we exclude the trivial cases $x = 0$ (where ε necessarily equals 0^∞) and $x = 1/(q-1)$ (where $\varepsilon = 1^\infty$). Secondly, if $q \leq G$, where $G = \frac{1+\sqrt{5}}{2} = 1.618\ldots$, then the set of such x's is known to be empty – see [7, Theorem 3]. So, we will consider only $q \in (G, 2)$ and $x \in (0, 1/(q-1))$.

If a sequence ε is a unique representation, it must be both the *greedy* ("choose 1 whenever you have a choice") and *lazy* ("choose 0 whenever you have a choice") expansion of x.

The characterization of both expansions is well known (see, e.g., [7, 15]). Namely, let the sequence $(d_n)_1^\infty$ be defined as follows: let $1 = \sum_1^\infty d'_k q^{-k}$ be the greedy expansion of 1, i.e, $d'_1 = [q] = 1, d'_2 = [q\{q\}]$, etc. If the tail of the sequence (d'_n) differs from 0^∞, then we put $d_n \equiv d'_n$. Otherwise let $k = \max\{j : d'_j > 0\}$, and $(d_1, d_2, \ldots) := (d'_1, \ldots, d'_{k-1}, d'_k - 1)^\infty$. In the seminal paper [15] it is shown that if $x < 1$ and (1) is the greedy expansion, then $(\varepsilon_n, \varepsilon_{n+1}, \ldots)$ is lexicographically less (notation: $\prec$) than $(d_1, d_2, \ldots)$. Recall that $(x_n)_1^N \prec (y_n)_1^N$ (where $1 \leq N \leq +\infty$) iff $x_n < y_n$ for the smallest n such that $x_n \neq y_n$. Similarly, if $x > \frac{2-q}{q-1} = \frac{1}{q-1} - 1$, then $(\bar{\varepsilon}_n, \bar{\varepsilon}_{n+1}, \ldots) \prec (d_1, d_2, \ldots)$, where the bar denotes the *reflection*, i.e., $\bar{1} = 0, \bar{0} = 1$.

It is natural to restrict x to the interval $\Delta_q = \left(\frac{2-q}{q-1}, 1\right)$; however, in the end we return to the interval $(0, 1/(q-1))$ – see Corollary 15. Let

$$\mathcal{A}_q := \{x \in \Delta_q : \text{the representation (1) is unique}\}.$$

Proposition 1. [6] *The set $\mathcal{A}_q$ has Lebesgue measure zero for any $q \in (G, 2)$.*

One might think that the Hausdorff dimension of $\mathcal{A}_q$ is also always zero but this is not the case for q close to 2. Let us formulate the main theorem of the present paper.

Theorem 2. *The cardinality of the set $\mathcal{A}_q$ is*

(i) *2, if $q \in (G, q_f]$, where q_f is the root of the equation*

$$x^3 = 2x^2 - x + 1, \quad q_f = 1.7548776\ldots;$$

(ii) $\aleph_0$ *if $q \in (q_f, q_c)$ and*
(iii) $2^{\aleph_0}$ *if $q \in [q_c, 2)$.*

Moreover, if $q \in (q_c, 2)$, then $\mathcal{A}_q$ is a Cantor set (i.e., it is perfect and nowhere dense), and $\dim_H(\mathcal{A}_q) > 0$.

Thus, for this problem the "Komornik-Loreti constant" q_c is critical as well – it separates a Cantor set of positive Hausdorff dimension from a countable set!

Remark 3. Note that formula (1) may be regarded as a map $\pi_q : \Sigma \to \mathbb{R}$ and by definition, it is one-to-one on the preimage of $\mathcal{A}_q$ which we will denote by $\mathcal{U}_q$. Theorem 2 (i) is implicitly contained in [5], where some partial results on the structure of $\mathcal{U}_q$ were also obtained; however, for the reader's convenience we are going to give the proof of this item as well (especially as it is very simple).

Our tool will be the following lemma which describes the set $\mathcal{U}_q$ (see above):

Lemma 4.
$$\mathcal{U}_q = \varepsilon \in \Sigma : (\varepsilon_n, \varepsilon_{n+1}, \dots) \prec (d_1, d_2, \dots)$$
$$and \ \ (\overline{\varepsilon}_n, \overline{\varepsilon}_{n+1}, \dots) \prec (d_1, d_2, \dots), n \in \mathbb{N}\}.$$

3. Proof: above the critical value

We will begin with the proof of Theorem 2 (iii). From Lemma 4 it follows that the set $\mathcal{U}_q$ is a non-decreasing function of q, i.e., if $q' \leq q$, then $\mathcal{U}_{q'} \subset \mathcal{U}_q$ (this is because the sequence (d_n) is an increasing function of q in the sense of the lexicographic order – see [15]) . Therefore, to prove Theorem 2 (iii), it suffices to construct a sequence $q_n \searrow q_c$ such that $\mathcal{U}_{q_n}$ is uncountable and $\dim_H(\pi_q(U_{q_n})) > 0$ for any $q \in (q_n, 2)$. We will call the sequences from $\mathcal{U}_q$ q-*admissible*. Let the notation $q : 1 \sim \varepsilon_1 \varepsilon_2 \dots$ indicate the fact that $(d_n) \equiv (\varepsilon_n)$, where $(d_n) = (d_n)(q)$ is as above.

We define q_n as the number for which $q_n : 1 \sim \mathfrak{m}_2 \mathfrak{m}_3 \dots \mathfrak{m}_{2^n} (\overline{\mathfrak{m}}_1 \dots \overline{\mathfrak{m}}_{2^n})^\infty$, i.e., $q_1 : 1 \sim 1(10)^\infty$, $q_2 : 1 \sim 110(1001)^\infty$, etc. It is obvious from (2) that $q_n \searrow q_c$, because $q_c : 1 \sim \mathfrak{m}_2 \mathfrak{m}_3 \dots$. Let σ denote the shift on sequences; following [12], we denote $\sigma(\mathfrak{m})$ by δ, i.e.,

(3) $\delta = 1101\ 0011\ 0010\ 1101\ 0010\dots$

We are going to prove a number of technical lemmas first. We will use the following relations: $\delta_{2^n} = 1$, $\delta_{2^n+j} = \overline{\delta}_j$ for $1 \leq j < 2^n$ (see [12]).

Lemma 5. *We have* $\delta_k \dots \delta_{k+2^n-2} \prec \delta_1 \dots \delta_{2^n-1}$ *for any* $n \geq 1$ *and any* $k \in \{2, \dots, 2^n - 1\}$.

Proof. By [12, Lemma 1], $\delta_k \delta_{k+1} \dots \ \prec \ \delta_1 \delta_2 \dots$, whence $\delta_k \dots \delta_{k+2^n-2} \ \preceq \ \delta_1 \dots \delta_{2^n-1}$. Assume that $n \geq 2$ and $\delta_k \dots \delta_{k+2^n-2} = \delta_1 \dots \delta_{2^n-1}$. Then $\delta_k = \delta_{k+1} = 1$, whence $\delta_{k-1} = 0$, and we can use the same argument as in the proof of [12, Lemma 1]. Namely, we have $k - 1 = 2^s + 2^m + j$, where $s \leq n - 1, m < s$ and $0 \leq j < 2^m$. It was shown in the proof in question that if $\delta_{k-1} = 0$, then $\delta_k \dots \delta_{k+2^{m+1}-j-1} \prec \delta_{j+1} \dots \delta_{2^m} \preceq \delta_1 \dots \delta_{2^m-j}$, whence $\delta_k \dots \delta_{k+2^n-2} \prec \delta_1 \dots \delta_{2^n-1}$, because $m + 1 < n$. □

Remark 6. In a similar way one can show that $\overline{\delta}_k \dots \overline{\delta}_{k+2^n-2} \prec \delta_1 \dots \delta_{2^n-1}$ for $k < 2^n$.

Lemma 7. *If* $\delta_k \dots \delta_{2^n-1} = \delta_1 \dots \delta_{2^n-k}$ *or* $\overline{\delta}_k \dots \overline{\delta}_{2^n-1} = \delta_1 \dots \delta_{2^n-k}$ *for* $k \leq 2^n - 1$, *then* $\delta_{2^n-k+1} = 1$.

Proof. We are going to prove both statements simultaneously using induction on n. For $n = 1, 2$ it is a direct inspection; let $n \geq 3$ and assume the claim to be true for any $i \leq n$ and $\delta_k \ldots \delta_{2^{n+1}-1} = \delta_1 \ldots \delta_{2^{n+1}-k}$. Note first that if $k < 2^n$, then $\delta_k \ldots \delta_{2^n+k-1} = \delta_1 \ldots \delta_{2^n}$, which contradicts Lemma 5. Hence $k > 2^n$ (it is obvious that $k \neq 2^n$); we have $\overline{\delta}_{k-2^n} \ldots \overline{\delta}_{2^n-1} = \delta_1 \ldots \delta_{2^{n+1}-k}$, and we can apply the induction hypothesis for $l = k - 2^n$. Hence $\delta_{2^{n+1}-k+1} = 1$. The case $\overline{\delta}_k \ldots \overline{\delta}_{2^{n+1}-1} = \delta_1 \ldots \delta_{2^{n+1}-k}$ is considered in the same way.   □

Lemma 8. *We have* $\overline{\delta}_k \ldots \overline{\delta}_{2^n-1}\delta_{2^n} \ldots \delta_{2^{n+1}-1} \prec \delta_1 \ldots \delta_{2^{n+1}-k}$ *for any* $n \geq 1$ *and any* $k \in \{1, \ldots, 2^n - 1\}$.

Proof. Again, by [12, Lemma 1], $\overline{\delta}_k \ldots \overline{\delta}_{2^n-1} \preceq \delta_1 \ldots \delta_{2^n-k}$. If this inequality is strict, we are done. Assume we have an equality here. Note first that by (3), $k \geq 5$; next, by Lemma 7, $\delta_{2^n-k+1} = 1$, and it suffices to show that $\delta_{2^n+1} \ldots \delta_{2^{n+1}-1} \prec \delta_{2^n-k+2} \ldots \delta_{2^{n+1}-k}$, which is equivalent to $\overline{\delta}_1 \ldots \overline{\delta}_{2^n-1} \prec \delta_{2^n-k+2} \ldots \delta_{2^{n+1}-k}$, which in turn is equivalent to $\overline{\delta}_{2^n-k+2} \ldots \overline{\delta}_{2^{n+1}-k} \prec \delta_1 \ldots \delta_{2^n-1}$. Since $2^n - k + 2 < 2^n$, the latter inequality follows from Remark 6.   □

Now we are ready to prove the main statement of this item. Let $v_n := \mathfrak{m}_1 \ldots \mathfrak{m}_{2^n}$.

Proposition 9. *If* $q > q_n$ *for some* $n \geq 2$, *then* $\mathcal{U}_q \supset V_n := \prod_1^\infty \{v_n, \overline{v}_n\}$.

Proof. We need to show that any 0-1 sequence which is constructed of blocks v_n and $\overline{v}_n$ is q-admissible for any $q > q_n$. Let $V_n' := \{\sigma^k \varepsilon : \varepsilon \in V_n, k \geq 0\}$. We claim that the maximal sequence in V_n' is $\sigma(v_n)(\overline{v}_n)^\infty$ – which is exactly the expansion of 1 for q_n. Let us prove this.

As v_n starts with 0 and $\overline{v}_n$ starts with 1, it is obvious that we must look for the largest element in V_n' among the sequences of the form $\sigma^k(v_n)(\overline{v}_n)^\infty$ or $\sigma^k(\overline{v}_n)(\overline{v}_n)^\infty$ for some $k < 2^n$. The cases $k = 0, 1$ are straightforward; let $k \geq 2$.

Let us first show that $\sigma^k(v_n)(\overline{v}_n)^\infty \prec \sigma(v_n)(\overline{v}_n)^\infty$. It suffices to prove the strict inequality for the segments of length 2^n, i.e. $\delta_k \ldots \delta_{k+2^n-1} \prec \delta_1 \ldots \delta_{2^n}$, which is exactly the claim of Lemma 5. Similarly, taking the segments of length $2^{n+1} - k$, we see that the inequality $\sigma^k(\overline{v}_n)(\overline{v}_n)^\infty \prec \sigma(v_n)(\overline{v}_n)^\infty$ follows from Lemma 8.

By the symmetry of V_n, the lexicographically smallest sequence in V_n' will be $\sigma(\overline{v}_n)v_n^\infty$. Both conditions of Lemma 4 are satisfied, whence $V_n \subset \mathcal{U}_q$.   □

Thus, $\mathcal{U}_q$ (and therefore, $\mathcal{A}_q$) has the cardinality of the continuum for $q > q_c$. To show that the Hausdorff dimension of $\mathcal{A}_q$ is positive, it suffices to prove that $\dim_H(\pi_q(V_n)) > 0$ for any $q > q_n$ (where V_n is as in Proposition 9). In fact, there is an explicit formula for the latter dimension, namely,

$$(4) \qquad \dim_H(\pi_q(V_n)) = \frac{\log 2}{2^n \log q} < 1.$$

This follows from the following simple lemma:

Lemma 10. *Let $Q > 2$ and π_Q be as above, i.e., $\pi_Q(\varepsilon) = \sum_{k=1}^{\infty} \varepsilon_k Q^{-k}$. Then*

$$\dim_H \left(\pi_Q(\prod_1^{\infty}\{a,b\}) \right) = \frac{\log 2}{\log Q}$$

for any $a \neq b$.

Proof. Let $a < b$; it is a direct inspection that $\pi_Q(\prod_1^{\infty}\{a,b\})$ is the middle-third Cantor set in $\left[\frac{a}{Q-1}, \frac{b}{Q-1} \right]$ with the middle part of length $\frac{(b-a)(Q-2)}{Q(Q-1)}$ and therefore it is homothetic to the middle-third Cantor set in $[0,1]$ with the middle part of length $1 - 2/Q$. The Hausdorff dimension of the latter set is known to be equal to $\log 2 / \log Q$ (the proof is essentially the same as for the classical middle-third Cantor set – see, e.g., [9, Chapter 1]). $\square$

To obtain (4), it suffices to apply this lemma to our framework with $q > q_n$, $Q := q^{2^n} > 2$ (as $n \geq 1$), $a := \sum_1^{2^n} \mathfrak{m}_i q^{-i}, b := \sum_1^{2^n} \overline{\mathfrak{m}}_i q^{-i}$. This proves that $\dim_H(\mathcal{A}_q) > 0$.

To conclude the proof of Theorem 2 for $q > q_c$, we need to show that $\mathcal{A}_q$ is a Cantor set. Note first that from Proposition 1 it follows that $\mathcal{A}_q$ is nowhere dense. Furthermore, if we endow Σ with its natural weak topology (the topology of coordinate-wise convergence), then the map $\pi_q : \Sigma \to \mathbb{R}$ is continuous. From Lemma 4 it is easy to deduce that $\mathcal{U}_q$ is closed, whence $\mathcal{A}_q$ is closed as well. Hence $\pi_q|_{\mathcal{U}_q}$ is a homeomorphism, as $\mathcal{U}_q$ is compact.

Since $\mathcal{A}_q$ and $\mathcal{U}_q$ are homeomorphic, it suffices to show that $\mathcal{U}_q$ does not contain isolated points. Let $q > q_n$ and assume first that $\varepsilon \in \mathcal{U}_q$ has a tail different from $(01)^{\infty}$. Then there exists an infinite number of indices k such that $\varepsilon_{k-2}\varepsilon_{k-1}\varepsilon_k = 110$ or 001. In view of the symmetry of $\mathcal{U}_q$ both cases are similar; let it be 110, for instance. Then it is easy to see that $(\varepsilon_1, \ldots, \varepsilon_k, v_n^{\infty})$ and $(\varepsilon_1, \ldots, \varepsilon_k, v_n(\overline{v}_n)^{\infty})$ are both q-admissible. This follows from the q-admissibility of ε and of $(1,1,0,v_n^{\infty})$. Therefore, ε cannot be an isolated point.

Assume now that $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_k, (01)^{\infty})$ is q-admissible. Then it is a direct inspection that $(\varepsilon_1, \ldots, \varepsilon_k, (01)^k, v_n^{\infty})$ and $(\varepsilon_1, \ldots, \varepsilon_k, (01)^k, (\overline{v}_n)^{\infty})$ are both q-admissible as well for any $k \geq 2$. Hence $\mathcal{A}_q$ is a Cantor set if $q > q_c$.

Remark 11. Note that actually there exists an explicit formula for the Hausdorff dimension of $\mathcal{A}_q$. Namely, let $u_n(q)$ denote the number of 0-1 words $(\varepsilon_1 \ldots \varepsilon_n)$, each of which can be extended to some sequence in $\mathcal{U}_q$. Note that by Lemma 4, the set $\mathcal{U}_q$ is obviously shift-invariant. Then the general theorem says that the topological entropy of the shift σ_q on $\mathcal{U}_q$ is given by the formula

$$(5) \qquad h_{\text{top}}(\sigma_q) = \lim_{n \to +\infty} \frac{\log u_n(q)}{n}$$

(see, e.g., [13]). Furthermore, by the standard result from the theory of cookie-cutters [10] and the fact that $\mathcal{A}_q = \pi_q(\mathcal{U}_q)$,

$$(6) \qquad \dim_H(\mathcal{A}_q) = \frac{h_{\text{top}}(\sigma_q)}{\log q} = \lim_{n \to +\infty} \frac{\log u_n(q)}{n \log q}.$$

One of the consequences of this formula is that the function $q \mapsto \dim_H(\mathcal{A}_q)$ is continuous (this is because the function $q \mapsto \{u_n(q)\}_{n=1}^\infty$ is continuous). However, it is not Hölder continuous, as $\dim_H(\mathcal{A}_q) \geq C/|\log(q-q_c)|$ when $q > q_c$ for some $C > 0$, because, as is easy to see, $q_n - q_c = O(q_c^{-2^n})$ and by Proposition 9 and formula (4), $\dim_H(\mathcal{A}_{q_n}) \geq 2^{-n}$.

4. Proof: below the critical value

Let us first prove Theorem 2 (i). Let $q \in (G, q_f]$ and suppose $\varepsilon \in \mathcal{U}_q$ and $\varepsilon_k = \varepsilon_{k+1} = 1$ for some $k \geq 1$. Since $q_f : 1 \sim (1100)^\infty$, we have by Lemma 4 and the monotonicity of $\mathcal{U}_q$, $(\varepsilon_{k+2}, \varepsilon_{k+3}, \dots) = ((0011)^\infty)$, which is not q-admissible, as $q \leq q_f$. In view of the symmetry of $\mathcal{U}_q$, the same argument works in the case $\varepsilon_k = \varepsilon_{k+1} = 0$. Therefore, there cannot be two consecutive equal symbols in $\mathcal{U}_q$, whence $(01)^\infty$ and $(10)^\infty$ are the only possible elements of $\mathcal{U}_q$. However, as $q > G$, both these sequences are indeed q-admissible, because $G : 1 \sim (10)^\infty$.

It remains to show that $\mathcal{U}_q$ is countable for $q \in (q_f, q_c)$ and has the cardinality of the continuum for $q = q_c$.

Lemma 12. *If $q \leq q_c$, then for any q-admissible sequence ε such that $\varepsilon_1 \dots \varepsilon_{2^n} = \delta_1 \dots \delta_{2^n}$ for some $n \geq 1$, necessarily $\varepsilon_{2^n+1} \dots \varepsilon_{2^{n+1}-1} = \delta_{2^n+1} \dots \delta_{2^{n+1}-1}$.*

Proof. By monotonicity, it suffices to consider the case $q = q_c$. Since ε is q_c-admissible and $q_c : 1 \sim \delta$, we have by Lemma 4, $\varepsilon_{2^n+1} \dots \varepsilon_{2^{n+1}-1} \preceq \delta_{2^n+1} \dots \delta_{2^{n+1}-1} = \overline{\delta}_1 \dots \overline{\delta}_{2^n-1}$. Hence $\delta_1 \dots \delta_{2^n-1} \preceq \overline{\varepsilon}_{2^n+1} \dots \overline{\varepsilon}_{2^{n+1}-1}$. At the same time, since ε is q_c-admissible, $\overline{\varepsilon}_{2^n+1} \dots \overline{\varepsilon}_{2^{n+1}-1} \preceq \delta_1 \dots \delta_{2^n-1}$, whence $\overline{\varepsilon}_{2^n+1} \dots \overline{\varepsilon}_{2^{n+1}-1} = \delta_1 \dots \delta_{2^n-1}$, and $\varepsilon_{2^n+1} \dots \varepsilon_{2^{n+1}-1} = \delta_{2^n+1} \dots \delta_{2^{n+1}-1}$. $\square$

Let $w_n := \sigma(v_n)1 = \delta_1 \dots \delta_{2^n}$, $n \geq 0$.

Proposition 13. *If $q_f < q < q_c$, then there exists $n = n(q)$ such that $\mathcal{U}_q$ contains only eventually periodic sequences with the period $w_k \overline{w}_k$ for some $k \leq n$.*

Proof. Suppose ε is q-admissible and $\varepsilon_m \dots \varepsilon_{m+2^k-1} = w_k$ or $\overline{w}_k$ for some $m \geq 1, k \geq 1$ (this always happens unless $\varepsilon = (10)^\infty$ or $(01)^\infty$). Let it be w_k, for instance. Then the next $2^k - 1$ symbols are determined by Lemma 12. For $\varepsilon_{m+2^{k+1}-1}$ there is a choice. If it equals 1, then $\varepsilon_m \dots \varepsilon_{m+2^{k+1}-1} = w_{k+1}$; otherwise $\varepsilon_m \dots \varepsilon_{m+2^{k+1}-1} = w_k \overline{w}_k$. Therefore, either we obtain the period $(w_k \overline{w}_k)$ or we will meet w_{k+1}. As q is strictly less than q_c, the latter event cannot occur for arbitrarily large k, because w_n cannot be the prefix of a q-admissible sequence starting with some n. Therefore, the periodic tail of the form described above, is inevitable for a q-admissible sequence if $q < q_c$. $\square$

Hence $\mathcal{A}_q$ is at most countable if $q < q_c$. However, as $q > q_f$, it is easy to construct a countable subset of $\mathcal{U}_q$; for example, take all the sequences of the form $(10)^k(1100)^\infty$ with $k \in \mathbb{N}$. Thus, card $\mathcal{A}_q = \aleph_0$, which proves Theorem 2 (ii).

Finally, from the proof of Proposition 13 and by monotonicity, $\mathcal{U}_{q_c}$ is exactly the set of all words of the form

$$(w_0\overline{w}_0)^{k_0} \ldots (w_n\overline{w}_n)^{k_n}, \quad k_j \in \mathbb{Z}_+$$

together with their shifts and reflections and therefore has the cardinality of the continuum. The main theorem is proven.

Remark 14. It is interesting to study the behaviour of the sequence $u_k(q)$. If $q > q_c$ then from (5) it follows that $u_k(q)$ grows exponentially at the rate $e^{kh_{\mathrm{top}}(\sigma_q)}$. For any $q < q_c$ it is easy to show that $u_k(q) = O(k^n)$, where $n = n(q)$ – this follows from the proof of Proposition 13.

The most intriguing case is $q = q_c$. Let $P(k)$ denote the number of representations of k as a sum $\sum_0^\infty e_j 2^j$ with $e_j \geq 0$. This combinatorial quantity is well studied; in particular, K. Mahler [14] showed that

$$\log P(k) \sim \frac{\log^2 k}{2\log 2}.$$

From our observation on the structure of $\mathcal{U}_{q_c}$ (see above) it easily follows that $u_n(q_c) \asymp P([n/2])$, whence

$$\log u_n(q_c) \asymp \log^2 n,$$

i.e., this sequence has an intermediate growth of order $n^{\log n}$.

Corollary 15. *The set $\mathcal{A}'_q$ of all $x \in (0, 1/(q-1))$ that have a unique representation of the form (1), is countable for $q \in (G, q_c)$ and uncountable with positive Hausdorff dimension for $q \in (q_c, 2)$.*

Proof. It suffices to note that every sequence ε' from $\mathcal{U}'_q = \pi_q^{-1}(\mathcal{A}'_q)$ is of the form $0^k\varepsilon$ or $1^k\varepsilon$, where $\varepsilon \in \mathcal{U}_q$ and $k \geq 0$. This follows, for instance, from the characterization of the greedy expansions for all $x \in (0, 1/(q-1))$ analogous to Parry's theorem. Namely, ε' is the greedy expansion of some $x \in (0, 1/(q-1))$ if and only if $(\varepsilon'_{n+1}, \varepsilon'_{n+2}, \ldots) \prec (d_1, d_2, \ldots)$ whenever $\varepsilon'_n = 0$ – see [8]. $\square$

Remark 16. Note that if $q < q_c$, then $\mathcal{A}_q$ is still closed but not perfect. Indeed, any sequence whose tail is $(w_n\overline{w}_n)^\infty$ for $n = n(q)$, is an isolated point. However, $\mathcal{A}_{q_c}$ is a Cantor set of zero Hausdorff dimension. We leave the proof of this claim to the reader as a simple exercise.

Example 17. Let $T = 1.839286\ldots$ be the "tribonacci number", i.e., the positive root of the equation $x^3 = x^2 + x + 1$. It lies above q_c, hence by Theorem 2, the set $\mathcal{U}_T$ must have the cardinality of the continuum. Let, as above, u_n denote the number of 0-1 words that can be extended to sequences in $\mathcal{U}_T$ – which is exactly the set of 0-1 sequences without three consecutive equal symbols. Hence it is easy to see that $u_1 = 2, u_2 = 4, u_3 = 6, \ldots$, and by induction, $u_n = 2F_{n+1}$, where $(F_n)_1^\infty$ is the Fibonacci sequence! Therefore, from the formula (6) it follows that

$$\dim_H(\mathcal{A}_T) = \lim_{n\to\infty} \frac{\log(2F_{n+1})}{n\log T} = \frac{\log G}{\log T} = 0.789677\ldots$$

A similar pattern occurs for the "multinacci" number T_n, i.e., the appropriate root of the equation $x^n = x^{n-1}+x^{n-2}+\cdots+x+1$, $n \geq 3$. Namely, $\dim_H(\mathcal{A}_{T_n}) = \frac{\log T_{n-1}}{\log T_n}$, whence it follows that $1-\dim_H(\mathcal{A}_q) \asymp 2-q$ as $q \nearrow 2$. For more details see [11].

Acknowledgement

Remark 18. During the preparation of this paper we learned from Jean-Paul Allouche that assertions (ii) and (iii) of Theorem 2 can also be deduced, using Lemma 4, from a result on the combinatorics of words and iterations of unimodal functions proved in his Thèse d'État [1].

References

[1] J.-P. Allouche, *Théorie des Nombres et Automates*, Thèse d'État, Université Bordeaux I, (1983).

[2] J.-P. Allouche and M. Cosnard, *The Komornik-Loreti constant is transcendental*, Amer. Math. Monthly **107** (2000), 448–449.

[3] J.-P. Allouche and M. Cosnard, *Non-integer bases, iteration of continuous functions, and an arithmetic self-similar set*, Acta Math. Hungar., to appear.

[4] J.-P. Allouche and J. Shallit, *The ubiquitous Prouhet-Thue-Morse sequence*, in C. Ding. T. Helleseth, and H. Niederreiter, eds., Sequences and Their Applications: Proceedings of SETA '98, Springer-Verlag, 1999, 1–16.

[5] Z. Daroczy and I. Katai, *Univoque sequences*, Publ. Math. Debrecen **42** (1993), no. 3–4, 397–407.

[6] P. Erdös and I. Joó, *On the number of expansions* $1 = \sum q^{-n_i}$, Ann. Univ. Sci. Budapest Eötvös Sect. Math. **35** (1992), 129–132.

[7] P. Erdös, I. Joó and V. Komornik, *Characterization of the unique expansions* $1 = \sum_{i=1}^{\infty} q^{-n_i}$ *and related problems*, Bull. Soc. Math. Fr. **118** (1990), 377–390.

[8] P. Erdös, I. Joó and V. Komornik, *On the number of q-expansions*, Ann. Univ. Sci. Budapest Eötvös Sect. Math. **37** (1994), 109–118.

[9] K. Falconer, *The Geometry of Fractal Sets,* Cambridge University Press, 1985.

[10] K. Falconer, *Techniques in Fractal Geometry,* John Wiley and Sons, 1997.

[11] P. Glendinning and N. Sidorov, *Uniqueness of β-expansions, a dynamical systems approach*, in preparation.

[12] V. Komornik and P. Loreti, *Unique developments in non-integer bases*, Amer. Math. Monthly **105** (1998), 636–639.

[13] D. Lind and B. Marcus, *An Introduction to Symbolic Dynamics and Coding,* Cambridge University Press, 1995.

[14] K. Mahler, *On a special functional equation*, J. London Math. Soc. **15** (1940), 115–123.

[15] W. Parry, *On the β-expansions of real numbers*, Acta Math. Acad. Sci. Hung. **11** (1960), 401–416.

[16] A. Rényi, *Representations for real numbers and their ergodic properties,* Acta Math. Acad. Sci. Hung. **8** (1957), 477-493.

Department of Mathematics, UMIST, P.O. Box 88, Manchester M60 1QD, United Kingdom.

E-mail address: `P.A.Glendinning@umist.ac.uk`

Department of Mathematics, UMIST, P.O. Box 88, Manchester M60 1QD, United Kingdom.

E-mail address: `Nikita.A.Sidorov@umist.ac.uk`